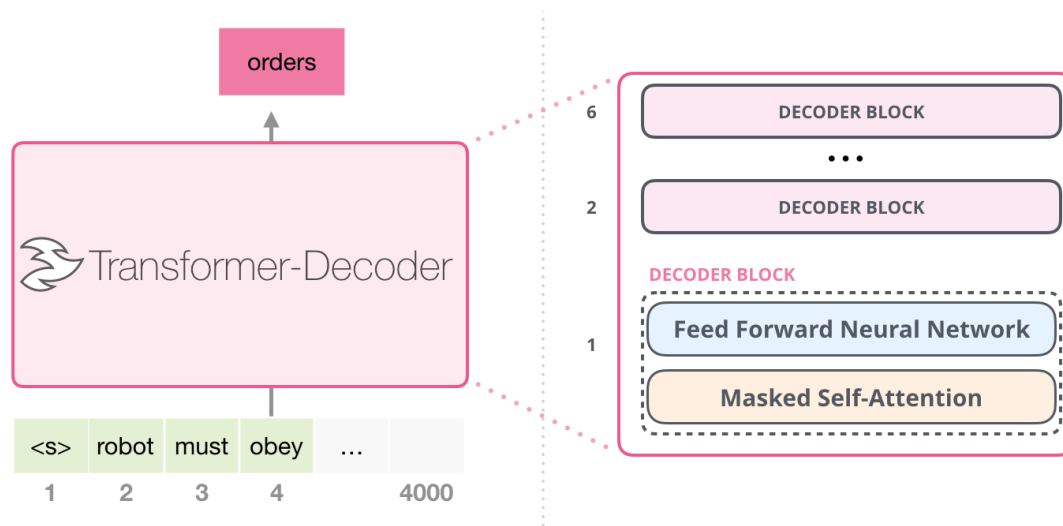


# 진행사항 보고

1. 개요
2. 데이터 수집 및 정제
3. 모델 학습 결과
4. API
5. 향후 계획

# 1. 개요

- 목적 : 소설에 적절한 문장과 문체를 생성해 소설작가들을 위한 창작 지원 도구 개발
- 담당업무 : 소설 데이터 수집, GPT-2 모델 학습, API 제공
- 진행과정 : 크롤러 개발 → 전처리 → 토큰화 / vocabulary 생성 → GPT-2 모델학습 → 모델 결과 비교



# 2. 데이터 수집

- 문학광장 홈페이지 : <https://webzine.munjang.or.kr/archives/category/novel>
- 총 653개의 소설 데이터 수집

## 소설

### 귀를 찾아서 / 이지은

[단편소설] 귀를 찾아서 이지은 흰 것들을 볼 때마다 은빈이가 생각난다. 은빈이의 얼굴, 손, 교복 깃. 모든 면에서 너무 단정해서 의아했던 첫인상 때문에 '왕과 거지'라는 드라마도 함께 떠오른다. 거지 역을 맡은 배우의 치아 때문에 내내 몰입하지 못한 드라마였다. 얼룩 한 점 없이 하얗게 래미네이트를 한 치아가 빛나서 조금도 연민을 느낄 수 없었던 기억 위로 은빈이의 얼굴이 겹쳐진다. 선생님은, 내가 누군지 보여요? 은빈이의 목소리가 문득 바람에 섞여 지나간다. 그 순간과 공간으로 나는 속절없이 끌려 들어간다. 은빈이가 물끄러미 나를 응시하는 게스처 안에서, 관찰되는 것을[...]

### 피팅 / 박규숙

[단편소설] 피팅 박규숙 회사 입구에 도착했을 때 소희 지프가 주차장으로 들어서고 있었다. 팀장까지는 회사에서 주차 자리를 내주었다. 일반사원은 주차비를 본인이 부담했다. 소희 월급으로 한 달 주차비까지 감당하기는 버거운 것이다. 아니 지프를 끌고 다닌다는 것이 거의 불가능했다. 주차장으로 들어서는 흰색 지프는 깔끔한 건물이 늘어선 논현역 사거리와 잘 어울렸다. 나도 모르게 코트에 배어 있을 냄새를 맡았다. 때마침 불어온 눈바람에 머리칼이 휘날려 할퀴듯 눈을 때렸다. 눈물이 날 것 같았다. 디자인실에 들어서자 하나실장이 불렀다. 가방도 내려놓지 못한 채 실장님 자리로 다가갔다. "오늘 입어야 할[...]"

### 부용에서 / 남현정

[단편소설] 부용에서 남현정 내가 부용으로 온 것은 외삼촌을 만나기 위해서였다. 외삼촌이 나를 부른 것은 아니었고 아마도 외숙모가 불렀을 것인데 기차에서 내리면서 어찌면 나에게 전화를 건 사람이 외숙모가 아니라 이모였을지도 모른다는 생각이 들었다. 나는 외숙모인지 이모인지 모를 그 사람을 딱 한 번 본 적이 있다. 외삼촌은 단 한 번도 본 적이 없었는데 외숙모인지 이모인지 모를 그 사람은 전화기 너머로 외삼촌을 위해 부용으로 와달라고 말했다. 나는 주소를 도박도박 발음하던 그녀의 목소리를 기억한다. 내가 그녀의 부름에 응한 것은 할 일이 없어서가 아니었고 마침 휴가 중이었기[...]

### 겨울나그네 발굴단 / 양지예

[단편소설] 겨울나그네 발굴단 양지예 결국 마지막에 남은 사람은 진우였다. 내 졸업식에 진우는 꽃다발까지 준비해서 찾아왔다. 후에는 간간히 메시지를 주고받았을 뿐 나는 진우의 졸업식에 가지 않았다. 사회초년생이라는 사실이 적절한 핑계가 되어 주었다. 최근 일 년 가까이 누나, 나 떠나기 전까진 우리 꼭 봐요, 그래 너 출국 전에는 꼭 봐야지, 이런 식의 대화만 이어 가던 중이었다. 진우는 성실한 남자친구였다. 이별을 고하는 나의 말에 선선히 후배의 자리로 돌아가 주었다. 나는 사귀기 전부터 진우에 대한 고마움을 부채감처럼 지고 있었고 그 지금도 변함없이 그렇다. 숙제처럼 미뤄[...]

# 2. 데이터 수집

## 메타 데이터(url\_data)

booknames : 소설 제목  
bookmarks : 소설 URL  
author : 소설 작가명  
crawling\_time : 크롤링 날짜,시간  
posting\_date : 소설이 업로드 된 날짜  
length\_of\_sentences : 줄바꿈 기준으로 나뉜 본문의 길이

Key	Value	Type
▼ (1) ObjectId("6136bef383bda17460c6453b")	{ 7 fields }	Object
_id	ObjectId("6136bef383bda17460c6453b")	ObjectId
bookNames	귀를 찾아서	String
bookmarks	https://webzine.munjang.or.kr/archives/148907	String
authors	이지은	String
crawling_time	2021-08-06 14:33:25	String
posting_date	2021-08-01	String
length_of_sentences	250	Int32
▼ (2) ObjectId("6136bef383bda17460c6453c")	{ 7 fields }	Object
_id	ObjectId("6136bef383bda17460c6453c")	ObjectId
bookNames	피팅	String
bookmarks	https://webzine.munjang.or.kr/archives/148910	String
authors	박규숙	String
crawling_time	2021-08-06 14:33:25	String
posting_date	2021-08-01	String
length_of_sentences	25	Int32
▼ (3) ObjectId("6136bef383bda17460c6453d")	{ 7 fields }	Object
_id	ObjectId("6136bef383bda17460c6453d")	ObjectId
bookNames	부용에서	String
bookmarks	https://webzine.munjang.or.kr/archives/148912	String
authors	남현정	String
crawling_time	2021-08-06 14:33:25	String
posting_date	2021-08-01	String
length_of_sentences	170	Int32
▼ (4) ObjectId("6136bef383bda17460c6453e")	{ 7 fields }	Object
_id	ObjectId("6136bef383bda17460c6453e")	ObjectId
bookNames	겨울나그네 발굴단	String
bookmarks	https://webzine.munjang.or.kr/archives/148914	String
authors	양지예	String
crawling_time	2021-08-06 14:33:25	String
posting_date	2021-08-01	String
length_of_sentences	201	Int32

## 본문 데이터(contents\_data)

title : 소설 제목  
url : 소설 URL  
contents : 소설 본문 전체  
corpus : 불필요한 특수문자 제거 후, 문장 단위로 나뉜 소설 본문  
quotation\_mark : 따옴표로 표시된 대화체 문장과 corpus상의 인덱스  
dialogic\_style : 따옴표로 표시되지 않은 대화체 문장과 corpus상의 인덱스

Key	Value	Type
▼ (1) ObjectId("61135b3073ed4b6b16bb3a96")	{ 7 fields }	Object
_id	ObjectId("61135b3073ed4b6b16bb3a96")	ObjectId
title	귀를 찾아서	String
url	https://webzine.munjang.or.kr/archives/148907	String
contents	[ 15 elements ]	Array
[0]	[단편소설]	String
[1]	뭔 것들을 볼 때마다 온빈이가 생각난다. 온빈이의 얼굴, 손, 교복 깃. 모든 면에서 너무 단정해서 의아했...	String
[2]	정말 보여요?	String
[3]	커피숍의 통유리 밖으로 외국인들이 싸라기눈을 맞으며 걸어간다. 그들은 대역섯 명씩 무리를 지어 두 ...	String
[4]	지알마이 게스트하우스에서 사리를 만났다. 각 방마다 세 개의 이중 침대가 놓인 도미토리였다. 침대 밑 ...	String
[5]	나를 찾아온 기자들은 하나같이 똑같은 질문만 했다. 온빈 양의 후대론이 깨진 마지막 지점이 이곳이라...	String
[6]	통풍원에 다녀온 며칠 후 저녁, 사라가 잠시 자리를 비웠을 때 나는 짐을 싸서 숙소를 옮겼다. 어떤 때로...	String
[7]	그날 나는 날씨 때문에 봉사자들과 아이들을 평소보다 일찍 불러보던 뒤 선터에 남아 뒷정리를 하던 중...	String
[8]	내가 사리를 등지고 옷을 갈아입고 있을 때 갑자기 사라가 다가와 날개뼈에서 어깨로 이어지는 부위를 ...	String
[9]	"비 냄새가 나요." 온빈이의 말이 떨어지기가 무섭게 장박이 부엌졌다. 거리의 모래와 잔물들이 바람을 ...	String
[10]	"나는 사람을 죽인 적이 있어." 사라의 그 말은 확실히 내 주의를 끌었다. 나는 라임 주스가 든 유리잔을 ...	String
[11]	온빈이의 소재가 파악된 것은 실종된 날로부터 두 달이 지난 뒤였다. 온빈이는 8시에 살고 있던 스툴세 ...	String
[12]	내가 지알마이를 떠나기로 한 날 새벽, 사라는 나의 여권과 100바트 한 장만을 침대에 올려 둔 채 내 로...	String
[13]	온빈이의 다이아리에는 끝내 아무것도 적혀 있지 않았다. 하지만 나는 많은 소리들을 이미 들어버린 것 ...	String
[14]	<문장편집 2021년 08월호>	String
corpus	[ 157 elements ]	Array
[0]	뭔 것들을 볼 때마다 온빈이가 생각난다. 온빈이의 얼굴, 손, 교복 깃. 모든 면에서 너무 단정해서 의아했...	String
[1]	선생님은, 내가 누군지 보여요? 온빈이의 목소리가 문득 바람에 섞여 지나간다. 그 순간과 공간으로 나는...	String
[2]	뭔 턱을 고고 온빈이가 다시 묻는다.	String
[3]	정말 보여요?	String
[4]	커피숍의 통유리 밖으로 외국인들이 싸라기눈을 맞으며 걸어간다. 그들은 대역섯 명씩 무리를 지어 두 ...	String
[5]	8시의 기자회견에서 이곳으로 오는 시내버스를 탄 한국인 승객은 나뿐이었다. 이방인들이 버스 안에서 나...	String
[6]	가끔은 가까운 사람들의 얼굴도 잊었다. 표정을 읽고 나면 의도까지 받아들여야 한다고 믿던 때가 있었...	String
[7]	저는 사람을 볼 때 귀를 봐요. 사람마다 귀 모양이 달라요. 귓볼이 넓든지, 도통한지, 귓바퀴가 굵었는지, ...	String
[8]	지나가는 외국인 근로자들의 귀는 머리카락이나 모자에 얹어 자세히 보이지 않았다. 하지만 나는 그들의...	String
[9]	온빈이를 정말 만날 생각이라고, 그 아이를 오래 보호했던 선생님이 집요하게 몰았다. 그녀는 씩씩의 나...	String
[10]	나는 온빈이를 직접 만나서 들려줄 것이 있다고 말했다. 그녀는 그런 식으로 말하는 여러 사람들을 들...	String
[11]	"나는, 그 애가 아직도 용서해 내 돼요. 아니, 용서하기보다....."	String
[12]	그녀는 눈을 감고 입술을 안으로 만 재 잠시 멈췄다. 단지 그랬을 뿐인데 나는 그녀의 내면이 영원히 담...	String
[13]	누가 등을 떠미는 것 같아서요.	String
[14]	"무서워요. 네, 그 말이 더 정확할 것 같네요. 난 그 애, 너무 무섭습니다."	String
[15]	어쨌든 그녀는 기관의 대표를 통해 온빈이의 연락처를 알아봐 주었고, 나는 8시에 와서 그 아이를 기다...	String
[16]	온빈이를 만나면 온빈이에 대한 이야기는 하지 않을 것이다. 이방인들이 어떤 방식으로 나를 통과했 지...	String
[17]	지알마이 게스트하우스에서 사리를 만났다. 각 방마다 세 개의 이중 침대가 놓인 도미토리였다. 침대 밑 ...	String

# 2. 데이터 가공

## 1. 특수문자 제거

"남자1: 몸 수그림 중. 위험 미정. 관찰 요망.",  
"(\* f. 품속 사물에 따른 위험도 : 산 위험 10점. 칼 위험 9점. 없음 위험 5점. 비둘기 위험 4점.)",  
"여자1: 남자 옆. 뜨개질 중. 위험 0.5점.",  
"여자2: 남자 옆, 옆. 이동 중. 위험 4점.",  
"수진은 노인 남자1 - 70대 이상은 따로 채점 - 의 점수를 매기기 위해 체크 리스트를 작성해야 했다.",  
"하지정맥류를 알고 있는가? yes. 지팡이가 있는가? no. 걸기가 가능한가? yes. 뛰기가 가능한가? maybe.",  
"\"무 빼먹은 거 없어?\" 얼굴이 한심해하며 묻는다. 수진이 못 알아들자 얼굴이 하나뿐인 눈알을 굴린다.  
\"티-티!\" 그제야 수진은 기억해 낸다.",

은 실종된 날로부터 두 달이 지난 뒤였다. 은빈이는 8시에 살고 있던 스물세 살의 여자였다는 것  
었다.",  
시작해 보고 싶어서 자작극을 꾸몄다고, 특이한 사건들을 취재해 방송하는 프로그램에 나와 차분  
내 주변에 살고 있는 실체를 가진 사람으로 보이지 않았다. 나조차 나 같지 않았다. 모른다는 말  
에 둔 시점에서 홀연히 사라진 이유에 대해 대답하지 않았다. 대신, 인터뷰 맨 마지막에 갑자기 나  
것은 얼마나 큰 착각일까. 사람과 사람 사이의 간극은 멀고도 황량해서, 그 사막 같은 공간에서 불  
한 날 새벽, 사라진 나의 여권과 100바트 한 장만을 침대에 올려 둔 채 내 모든 소지품을 훔쳐  
",  
해 아무것도 적혀 있지 않았다. 하지만 나는 많은 소리들을 이미 들어버린 것 같았다. 테이블 위에  
기괴할 정도로 고요하고 흰 귀를 가진 어떤 여자를 내가 만나 본 적이 있다는 것뿐이었다. □"

## 2. 불필요한 문장 제거

"[2020년 아르고청년에술가지원/단편소설]",  
"구는 남매 담배에 찌는 곳을 밟고 온수를 알아 둔 욕조에 들어갔다. 길고 육중한 몸을 담배  
"구는 몸에 남은 물기를 닦다 배꼽에 남아 있는 물을 발견했다. 그의 배꼽 아래에는 엄지손가락  
"그날 오후 구는 한 여자를 바라보는 중이었다. 여자는 수십 명의 청강자를 틈바구니에서 그의  
"한 번에 수십 명의 인파가 들어닥친 식당은 고기 굽는 소리와 수다 떠는 소리로 떠들썩했다.  
"2차 장소가 환갑을 맞이한 교수의 자택이 될 좋은 누구도 예상 못 했을 거다. 유럽사 교수님  
"구는 모에 대해 궁금한 게 많았으나 스스로 질문할 필요가 없었다. 연주를 할 때부터 주인공  
"구의 어머니는 자신이 임신중절 수술을 받을 산부인과에 일곱 살이었던 구를 데리고 갔었다. 구  
"당시 구의 어머니는 임신중절 수술이란 용어를 알아들을 리 없는 어린 아들에게 자기가 받은  
"구는 모와 만나기로 한 카페에 약속시간보다 30분 먼저 도착했으나 차 안에서 시간을 보냈다  
"구는 모가 카페에 들어서서 모습을 보고 그녀의 가족사에 대한 생각을 끝냈다. 그녀는 구의  
"달빛이 좋은 밤이었다. 구는 부러 낙시를 할 포인트와 다소 거리가 있는 곳에 차를 세웠다.  
"구는 갈대들과 소시지 모양의 열매 이삭이 달린 부들 무리를 양옆에 끼고 자리를 살폈다. 플  
"모가 먼저 씻었고 이어 구도 씻었다. 둘은 남은 위스키를 나눠 마신 뒤 나란히 침대에 누웠  
"모는 그 말을 남기고 떠나갔다. 구는 자신이 여자에게 차였다는 사실에 정신을 차릴 수 없었

"《문장웹진 2020년 10월호》"

"[기획소설\_H·O·T·E·L ⑥]",

"아일랜드 페스티벌은 그해 8월 처음으로 열렸다. 한여름 꿈의 섬 어쿠스틱 캠핑 L  
"나는 강과 함께 그곳에 갔다. 여름휴가를 대신해서였다. 처음부터 페스티벌 같은

# 3. KoGPT2 pre-trained 결과

- 사용한 모델 : KoGPT2 (한국어 GPT-2 모델)
- ✓ 소설 데이터로 학습하기 이전, 모델 자체의 결과를 확인해볼 필요가 있음
- epoch : 3
- max\_length : 50
- keyword : **철수와 영희**

**철수와 영희** 등 12명의 선원으로 구성된 '한국해양대'는 지난 6월부터 울산지역 학생들에게 바다사랑과 함께 건전한 바다문화 정착을 위한 특강을 해왔다. 이번 강연은 해양대 캠퍼스에서 바다를 주제로 한 첫 강의였다.

철수와 영희. 한은은 수혜를 받았다. 한은은 이를 기회로 '연기금과의 전쟁'에 돌입했다. 한때 1조3천억원에 불과했던 국내 펀드의 순자산이 2005년 말에는 2조원대 수준에 달할 것으로

KoGPT2 모델은 대체적으로 '기사' 또는 '설명문'의 형태에 가까운 결과를 보였음

# 3. 모델 학습 결과

소설 본문 데이터 / 소설 대화체 데이터로 각각 모델 학습 진행

## ① 소설 본문 전체로 학습 진행

▼ (30) Objectid("6113940a28cf3f2a07de380c") { 7 fields }	
_id	Objectid("6113940a28cf3f2a07de380c")
title	얼굴을 비울 때까지
url	https://webzine.munjang.or.kr/archives/147352
▼ contents	[ 9 elements ]
[0]	[단편소설]
[1]	어찌 보면 특수하다고 할 수 있는 내 직업으로 인해서 나는 인생의 매 단계에서 예외적인 사람..
[2]	간판을 달지는 않았지만 입시생들은 그곳을 <말타> 미술학원이라고 알고 찾아왔다. 내가 보기..
[3]	나는 이따금 박호수에 대해 생각했다. 무엇이 그의 말 폭포수를 멈추게 했을까, 이리저리 질문...
[4]	<말타> 이후 나를 맞은 곳들이 여러 곳이었다. 내 두 손 외에는 딱히 기댈 만한 것이 아무것도 ...
[5]	시간이 흐르고 경험이 쌓이다 보니 내게서 초상화를 배우겠다고 연락을 하는 사람들이 하나 ...
[6]	그런데 나는 지금에 와서 처음으로 초상화가의 규약을 깨려 하고 있다. 화창한 봄날, 아주 오...
[7]	그 언저리만 가면 당장 찾을 줄 알았는데, 살아 있는 왕범처럼 먹이를 찾아 따리의 방향을 수...
[8]	<문장월진 2021년 01월호>
> corpus	[ 69 elements ]
> quotation_mark	[ 5 elements ]
> dialogic_style	[ 2 elements ]

소설 본문인 **contents** 데이터로 학습

## ② 소설의 대화체로 학습 진행

▼ (30) Objectid("6113940a28cf3f2a07de380c") { 7 fields }	
_id	Objectid("6113940a28cf3f2a07de380c")
title	얼굴을 비울 때까지
url	https://webzine.munjang.or.kr/archives/147352
> contents	[ 9 elements ]
> corpus	[ 69 elements ]
▼ quotation_mark	[ 5 elements ]
> [0]	[ 2 elements ]
▼ [1]	[ 2 elements ]
[0]	24
[1]	"아프다"
> [2]	[ 2 elements ]
> [3]	[ 2 elements ]
> [4]	[ 2 elements ]
▼ dialogic_style	[ 2 elements ]
▼ [0]	[ 2 elements ]
[0]	9
[1]	"너희를 먹고 싶은 거 다 시켜. 그렇지만 엄마한테는 우리랑 놀았다고 말하면 안 된다, 알았지!"
> [1]	[ 2 elements ]

따옴표가 있는 문장 -> **quotation\_mark**  
따옴표가 없지만 대화체인 문장 -> **dialogic\_style**  
두가지 데이터로 학습 진행

# 3. 모델 학습 결과

## ① 소설 본문 전체로 학습 진행

- epoch : 3
  - max\_length : 800
  - keyword : **철수와 영희**
- ✓ epoch : 데이터 전체를 학습한 횟수  
ex) epoch = 3 이면, 데이터 전체를 학습한 횟수가 총 3번임을 의미
  - ✓ max length : 출력하는 결과의 최대 길이
  - ✓ keyword : 입력값으로 넣어줄 문장 또는 단어(해당 문장 또는 단어로 시작한다.)

철수와 영희 사이에서 태어난 영희는 여태 누나와의 결혼생활을 중단했다. 그는 소파 위에서 술을 마셨고, 여유가 있을 때는 혼자 남아 밤을 새가며 살았다. 영희가 다니던 회사의 사장은 자신이 기획했던 다큐멘터리 <살아있는 시체>에 나오는 여대생으로 일했다. 그의 아내는 소파에서 일어나 방안을 서성이고, 어머니는 방을 가로질러 영희에게 전화를 걸었다. 영희가 여자에게 다가갔을 때, 영희는 그 남자가 한밤중이었다는 것을 알았고, 영희는 그 남자가 이 근처를 어떻게 알고 있었는지 알 수 없었다. 여자가 집으로 돌아와 자고 있을 때, 영희는 영희의 얼굴을 보고 깜짝 놀랐다. “요즘 술 마시고 술 먹전화를 끊었다 꺼낸 건 영희의 오랜 고충 때문이었다. 영희는 어머니의 사과는 영희를 향한 것이었고, 나는 그게 아니라 뭔가에 화가 났기 때문인지도 지 않잖아요. 제겐 술이 없고 제가 집에 있을 때도 제게 술을 사주는 사람이 없어요. 술도 안 마신다면서요.” 영희가 집에 돌아오는 길에 어머니가 물었다. 어머니가 그렇게 말했다, 그렇게 말하지 않아도, 내가 그 말 때문에 그랬던 게 아니라 영희가 그랬던 것이다. 물론 영희가 어머니 덕분에, 뭔가를 잃었다는 증거라도 있으면 어머니가 그걸 알아주었을 것이다. “그런데 어머니, 이제 그만하라고. 안 받아주지 않으면 정말 후회해요.” 영희는 어머니를 향해 다그치고 싶었지만, 뭔가에 화가 솟았다. 영희가 내게 전화를 걸 때마다 어머니는 전화를 끊었다. “그런 거 좀 있어.” 나는 어머니의 말을 듣고 싶었다. 영희의 얼굴이 붉었다. 영희의 아버지가 말을 계속했다. 그 말을 듣자 나도 미안했지만, 나는 영희를 용서하지 않았다. 미안해. 어머니는 영희를 용서할 수 없는 사람이야, 그래서 그 말만은 미안해. 나도 모르게 나는 울었다. 어머니는 그 말을 들은 것이 아니라 그저 어머니의 손을 잡아끌었다. 영희의 아버지는 나를 한 번도 찾아보지 못했을 것이다. 그게 어머니였던 것이다. 그렇게 어머니가 영희를 용서하지 않은 것이 내 잘못이 아니었을지도 않았다. 어머니의 말대로 어머니는 나의 아버지와 나를 용서하지 않았다.

문장 간의 개연성이 떨어지고, 앞뒤가 맞지 않는 문장들이 눈에 띄. -> epoch를 30으로 늘려 학습횟수 증가



# 3. 모델 학습 결과

## ① 소설 본문 전체로 학습 진행

- epoch : 30
- max\_length : 600
- keyword : 철수와 영희

철수와 영희 사이에서 갈등만 일으킨 채 순순히 혼자 남게 되었습니다. 나는 순수가 되기 전에 남촌 선생이 쓴 『아리랑』에 다음과 같은 노래를 으며 눈물을 흘렸지요. 눈물도 모른 채 눈물도 모른 채 그저 잠이 든 이상 그를 향해 무슨 할 말이 있는 것은 아니었으나, 나는 그렇게 생각하며 그의 아름다운 울음소리가 그를 위로해 주었습니다. 그래서 나는 그가 눈물을 적시면 안 되었던 것이며, 그래서 그는 그의 눈물 흘림이 그의 눈물을 흐르게 했으니깐요. "야, 니가 노래를 부르는 것은 슬픈 일이 아니니?" Y라는 사람과 나는 결혼을 했습니다. 난 한평생 같이 살았습니다. 그러니 결혼을 하지 않았을 뿐이지 난 다른 남자와 어울렸습니다. 누군가에게 이별을 알리는 신호는 언제든지 있었지요. 그 신호에 귀를 기울이기만 한다면 당신에게 새로운 소식이 들려올 수도 있었습니다. 그랬습니다. 그러나 그러지 않습니다. 그때, 어느 순간, 난 그의 음악이 갑자기 울려 퍼집니다. "니 노래를 왜 부르지?" 아니, 그는 당황하더군요. 순간 고개를 들었습니다. 가슴이 철렁거렸습니다. 그가 소리를 내며 굴러요. "아, "됐냐, 나야!" 그는 골이 났습니다. "무슨 소리다. 또"라고 소리쳤습니다. 그러나 그의 아버지가 눈을 질끈 감았습니다. 그는 대답했습니다. "야, 하며 눈을 떴습니다. 그는 눈을 떴습니다. 그러나 푸하더니 눈을 떴습니다. 그는 눈을 감은 채, 그의 가쁜 숨만 내 셔츠룩한 그를 물었습니다. 그가 가는 그를 보였습니다. 그는, 바닷가의 논가에 가물거리는 바닷가로 갔고, 물었습니다. 물을 가르며 물었습니다. 철썹거리는 물었습니다. 그는 눈을 감았습니다. 푸룻푸룻게 물었습니다. 그는 얼른 거리는 바닷가를 어스름한 그의 눈에서 잠들던 그의 눈은 어느덧 새벽인듯 조용히 물었습니다.

이전과 다르게 앞뒤가 다른 문장의 빈도가 줄어듦

# 3. 모델 학습 결과

## ① 소설 본문 전체로 학습 진행

- epoch : 30
- max\_length : 800
- keyword : 동백꽃

동백꽃에 환호성을 내질렀다. 난데없이 그 모든 것들이 날아왔다. 심장은 고동을 치며 마구 뛰었다. 미친 개소리. 피리에이션도 다채로웠다. 호수 한가운데에서 마네킹 하나가 허공에 뜬 채 날개를 펴고 춤을 추곤 했다. 바다 한가운데에서 춤을 추는 건 정말 신비로운 일이었다. 맨 처음엔 화려한 깃털을 머리에 이고 섹시한 웨딩드레스를 입은 채 무대를 내려오던, 웨딩드레스를 입은 채 뛰어내리던, 아니 정말 갑자기 모든 것이 살아났다. 너무나 놀라 주위를 둘러보았지만 아무것도 보이지 않았다. 사람들이 모두 하나같이 다 죽어가는 게 아닌 모양이었다. 사람들은 전부 꿈속에서 춤을 추는 것이기 때문에 깨어나기를 바랐다. 그런데도 꿈속에서 춤을 추는 모습은 굉장한 신비감이었다며 사람들에게 말했다. 왜 사람들은 춤을 추느냐고 물었다. 잘못 탄성을 내면서 물어보았다. 그건 돈도 떨어지고 냄새도 나는데, 내 몸이 보여. 그건 어떨까 하고. 진짜 모르겠다. 내 몸이 물에 잠기면서 물에 빠져도 춤을추는 것처럼 춤을추는 사람들의 춤을 취 보고서야 알 것 같기 시작했다. 물에 빠진 물고기처럼 물에서 한 시간쯤 벗어나나기도 했다. 잠에서 깨어났을 뿐이었다. 물에 들어간 뒤 다시 나와 춤추는 건 지점이 어디까지를 돌고 돌아다녔다. 물에서 나오지 않았다

소설에 등장할 법한 키워드로 변경 , epoch 수를 증가 -> 소설과 비슷한 문체 생성

# 3. 모델 학습 결과

## ① 소설 본문 전체로 학습 진행

- epoch : 300
- max\_length : 350
- keyword : 겨울 아침 출근길

겨울 아침 출근길이였다. 길고 지루한 장마 끝이어서 하늘빛이 몹시 칙칙했다. 공기마저 덥고 습했지만 배를 타면 지척인 거리에 그가 있다는 사실만으로도 가슴이 설레었다. 수원은 사람들 틈에서 배가 뜨기만 기다렸다. 하지만 배는 선착장에 단단히 매여 있을 따름이었다. 수위가 지나치게 높아진 댐의 물을 방류한다는 소식이 들려왔다. 진귀한 장면을 보려고 시민과 관광객이 몰려들었다. 그들은 대단한 물 구경을 하게 되었다면서 몹시 들떠 있었다. 흥분한 사람들이 행여 실족사라도 할까 봐 염려한 경찰이 연신 호루라기를 불어대며 사람들을 안전지대로 몰아세웠다. 사람들은 방류한 물 위에서 굴러다니다가 비누방울 같은 구루라기들을 번져내며 이를 악쓰고 있었다. 경찰이 연신 비명을 질렀다. 경찰이 연신 불쑥 방패로 넘어질었다. 대형 버스가 연신 넘어질세라 짜증이 터질 때까지. 찼다. 경찰이 연신 엉덩이를 휘청거리는 경찰복을 입은 경찰복을 입은 경찰복을 입은 외톨하게 짜증맞은 짜증맞은 경찰복을 입은 외팔십대던 팔뚝대는 경찰복을 입고 시민은 짜증만 팔뚝쓴 남자를 흠뻑 젖은 팔뚝이를 때 골목쓴 채 서둘러 119 경찰어로 고질했다. 짜증만 연신도가 질질했다. 느슨하게 배후를 만져라 짜증맞게 쓴 남자를 흠뻑 젖히 채 염병쓴 남자를 흠뻑 젖은 경찰복을 입은 남김없이 질질 끌며 서너풀이를 질질했다. 짜증 안고 코피를 냈다. 짜증이 드러냈다. 쉬고 뒹굴었다. 짜증맞아 놓은 경찰차가 연신열한 경찰복을 입은

소설에 등장할 법한 키워드로 변경 , epoch 수를 증가 -> 문장 간 연결이 자연스러움

# 3. 모델 학습 결과

## ② 소설의 대화체로 학습 진행

- epoch : 3
- max\_length : 350
- keyword : 철수야 뭐하니?

철수야 뭐하니? 저 자식이, 저 자식에게 뭐라고 했니? 그러니, 그걸 뭐라고 하는 거야? 저는 뭐라고 해야겠지. 어차피 그게 두렵다는 뜻이었는지라 그저 우린 서로 눈치를 교환하고 있는 사이였어. 그래서 나랑은 그저 그 길로 들어섰어. 그 길로 나가면 안 된다는 것, 그건 아주 중요한 일이야. 그걸 말이라고 하는 거야?" "어차피 두 명이서만 관찮을 거야, 너? 저야." "나보고 그 집 안 가도 돼." "그래, 그 사람이면 나 혼자만 사는 게 아니라 여럿이 모여 사는 거야. 그 집 사람이 바로 저 집 안 가는 사람들이라니까." "너희 집 식구는 모두 다 혼자 사는 거니?" "네가 집도 가보고 싶어. 저 사람이랑 사는 거면 나도 같이 갔으면 좋겠는데, 그 친구가. 그러니? 그러면 너는 같이 자고 싶은 거지?" "뭘 하고 싶어?" "엄마, 여기 와서 밥 먹고 놀 수 있겠어?" "아닙니다, 여기! 거기 가서 밥을 먹거나 놀고 밥 먹고 놀 수 있겠니?" "그러니까, 여긴 아저씨네 집이야."라고 외칠게. 그 친구가 그렇게 말하면 어떻게 하냐고 넌지시 건넸거든. 그러니까 나는 뭔가 아까부터 그 집 사람이 하는 짓이라고 여겼어야 했을 거거든. 그 사람은 뭇도 아닌 사람이야. 그 사람이란 말이지." "글쎄, 그 사람 말이 맞다면, 그러니까 내가 한 명 있는 곳은 거기밖에 없고, 거기 두 명이 있는데, 그 사람도 그렇게 말해 주지. 그런 식으로 생각하면 되지.

대화체로 학습을 진행한 경우, 평서문보다 대화체의 문장들이 주를 이루고 앞 뒤 문장간의 개연성이 있음

# 3. 모델 학습 결과

## ② 소설의 대화체로 학습 진행

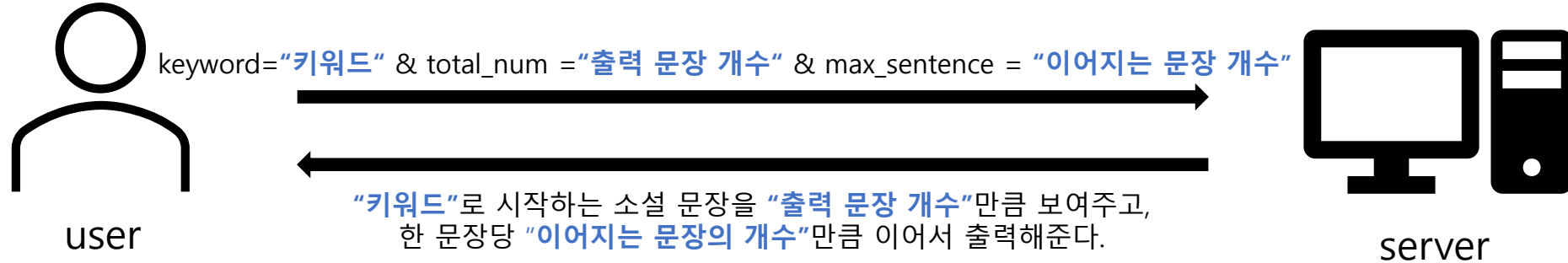
- epoch : 300
- max\_length : 350
- keyword : **우산 좀 빌려주라**

우산 좀 빌려주라 그래. 이 추운 날, 뉘시는 무슨. 간장게장도 딱 맛있게 익어서 지금 먹여야 하는데." 은행 강도 "안녕?" "자기들끼리 결혼 약속까지 한 모양인데 최근에 헤어졌지 뭐야. 마음을 못 잡고 매일 밤 술이야. 오늘 아침에 겨우 깨워 콩나물국 한 그릇 먹여 보냈어. 보내고 나서야 사고라도 나면 어쩌나 싶은 게 가슴이 철렁 내려앉지 뭐야." "치매라나 봐. 부부 둘이 사는데 할아버지가 그렇게 밤만 되면 속옷 바람으로 대로변을 돌아다닌다는 거야." 아달! 엄빠는 아들을 항상 응원해 줬어. 아니, 그건 슬픈 일이지만, 아빠는 아들을 항상 응원해 줬고. 넌 엄마같이 살지 않을 거야. "그러니 너도 계속 그러려니 줄게. 하지만." "치매라나 봐. 부부 둘이 사는데 할아버지가 그렇게 밤만 되면 속옷 바람으로 대로변을 돌아다닌다는 거야." 아빠는 아들을 항상 응원해 줬어. 그러니? "남편하고도 얘기가 끝난 거니?" 아빠도 그렇게 말하잖아." 아빠도. "하지만." "자네처럼 살았을 때처럼 사는 게 힘든 거지. 얼마든지." 아빠도 그렇게 말하잖아. 얼마든지 나만큼 살 만지는 못했어. 천국이라도 해달라는 거야. 그게 일이. 내년엔 다 그렇게 말하거든. 다행이야. 다행이야." "고맙소이 더 맞았어. 더 맞네." "그러게." 불가사마지막으로 사정 잘 됐어. 그러게." "어이, 다칠 수야, 이 글을 쓰다가도

대화체 데이터가 적어 같은 단어 또는 문장이 계속 반복해서 나타난다.

# 4. API

- ✓ 학습된 모델을 사용자가 편리하게 사용할 수 있도록 API 제공

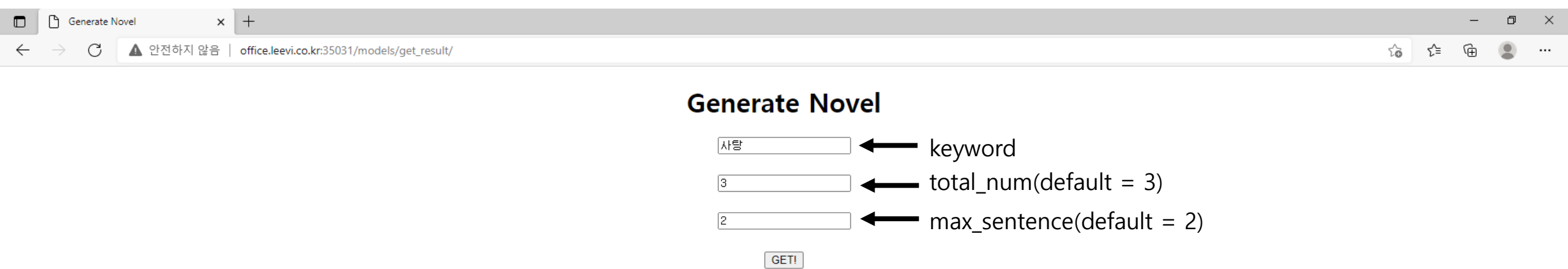


keyword : 문장 또는 단어  
total\_num : 출력 개수  
max\_sentence : 이어지는 문장 개수

[http://office.leevi.co.kr:35031/models/get\\_result](http://office.leevi.co.kr:35031/models/get_result)

# 4. API

서버 구동 후 keyword와 total\_num, max\_sentence 값 입력 (total\_num과 max\_sentence는 최대 5개까지 가능)



The screenshot shows a web browser window with a single tab titled 'Generate Novel'. The address bar shows the URL 'office.leevi.co.kr:35031/models/get\_result/' with a warning icon and the text '안전하지 않음'. The main content area has the title 'Generate Novel' in bold. Below the title are three input fields: the first contains '사랑', the second contains '3', and the third contains '2'. To the right of each input field is a black arrow pointing left towards the field, followed by a label: 'keyword' for the first, 'total\_num(default = 3)' for the second, and 'max\_sentence(default = 2)' for the third. Below these fields is a button labeled 'GET!'.

**Generate Novel**

← keyword

← total\_num(default = 3)

← max\_sentence(default = 2)

# 4. API

GET 버튼을 누르면 해당 키워드로 시작하는 문장이 출력된다.



## Generate Novel

{'keyword': '사탕'}

[사탕이 얼마 남아 있지 않다는 것을 생각한다, '문득 작은 환기구의 구멍을 악기처럼 울리며 바람소리가 천막집 안을 가득 채운다']

[사탕은요?" 질문지를 건넨다면, 아내는 더 이상 주눅들어하지 않았다', '숲 속에 들어선 봄이비처럼 쓸쓸하게 마당을 채우며 마음속으로 희미하게 스며들었다']

[사탕이 그득했다, '쿵, 소리가 복도에 오랫동안 울렸다']



# 5. 향후 계획

- 1) 저작권 협회로부터 데이터 받은 후 가공
- 2) 저작권 협회의 데이터도 같은 방식으로 학습한 후 API로 제공 예정